

Objective of the Project

Customer churn is a big problem for telecommunications companies. Indeed, their annual churn rates are usually higher than 10%. For that reason, they develop strategies to keep as many clients as possible. This is a classification project since the variable to be predicted is binary (churn or loyal customer). The goal here is to model churn probability, conditioned on the customer features.

Dataset

```
dataset = pd.read_csv('/content/telecommunications_churn.csv')
dataset
```

	account_length	voice_mail_plan	voice_mail_messages	day_mins	evening_mins	night_mins	international_mins	customer_service_calls	international_plan	day_calls	day_charge	ev
0	128	1	25	265.1	197.4	244.7	10.0	1	0	110	45.07	
1	107	1	26	161.6	195.5	254.4	13.7	1	0	123	27.47	
2	137	0	0	243.4	121.2	162.6	12.2	0	0	114	41.38	
3	84	0	0	299.4	61.9	196.9	6.6	2	1	71	50.90	
4	75	0	0	166.7	148.3	186.9	10.1	3	1	113	28.34	
...
3328	192	1	36	156.2	215.5	279.1	9.9	2	0	77	26.55	
3329	68	0	0	231.1	153.4	191.3	9.6	3	0	57	39.29	
3330	28	0	0	180.8	288.8	191.9	14.1	2	0	109	30.74	
3331	184	0	0	213.8	159.6	139.2	5.0	2	1	105	36.35	
3332	74	1	25	234.4	265.9	241.4	13.7	0	0	113	39.85	

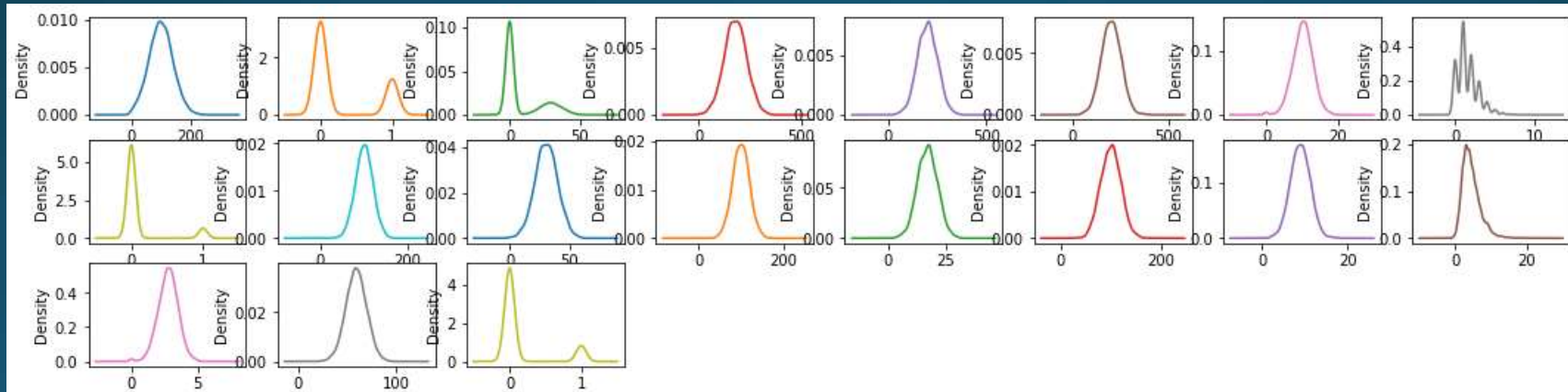
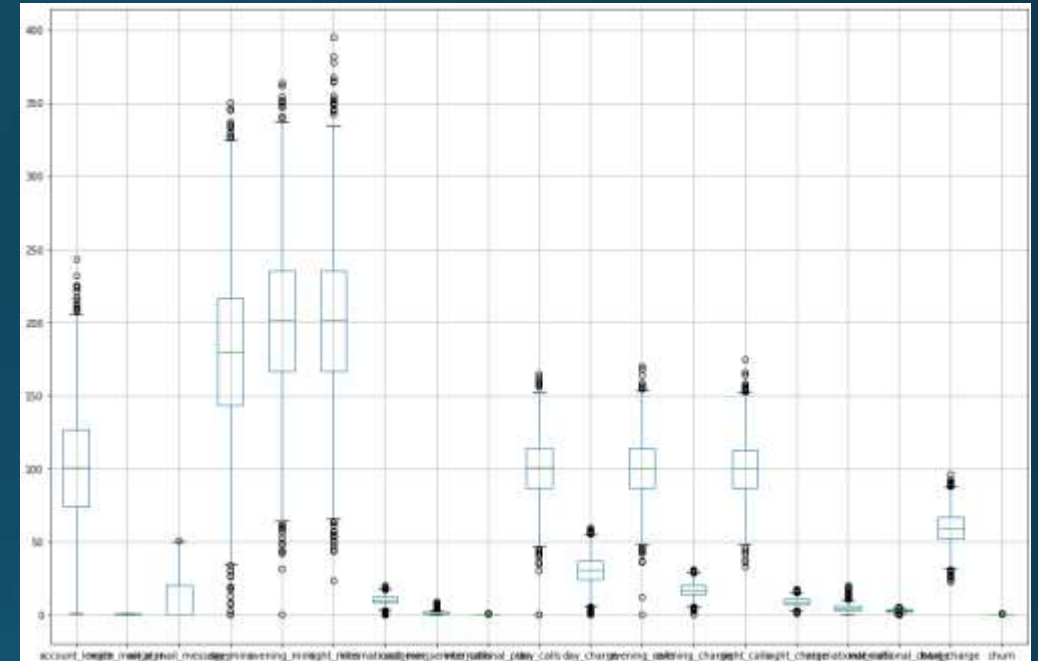
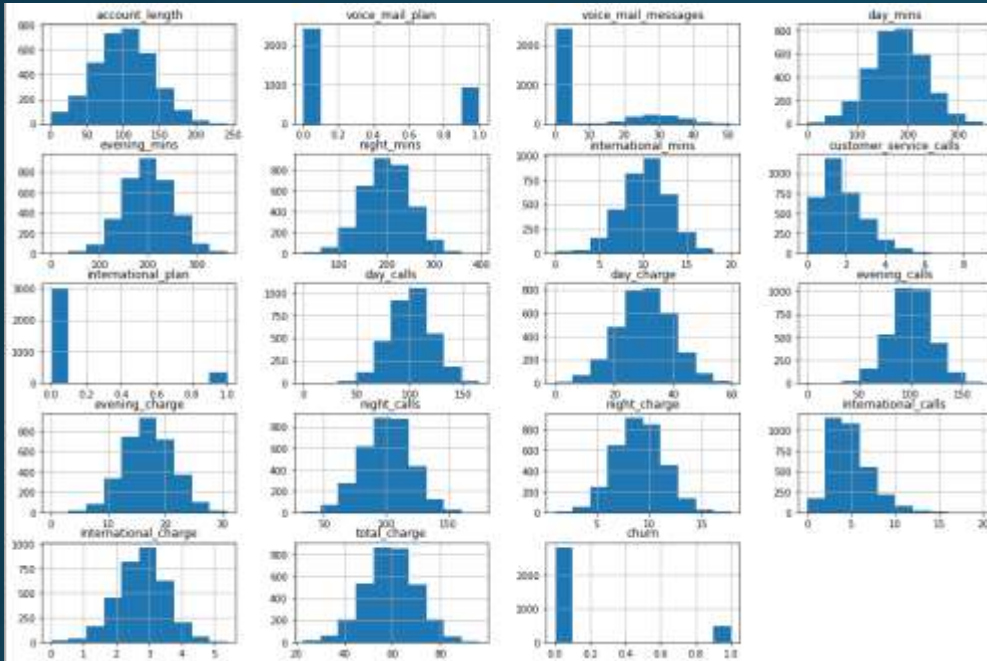
3333 rows × 19 columns



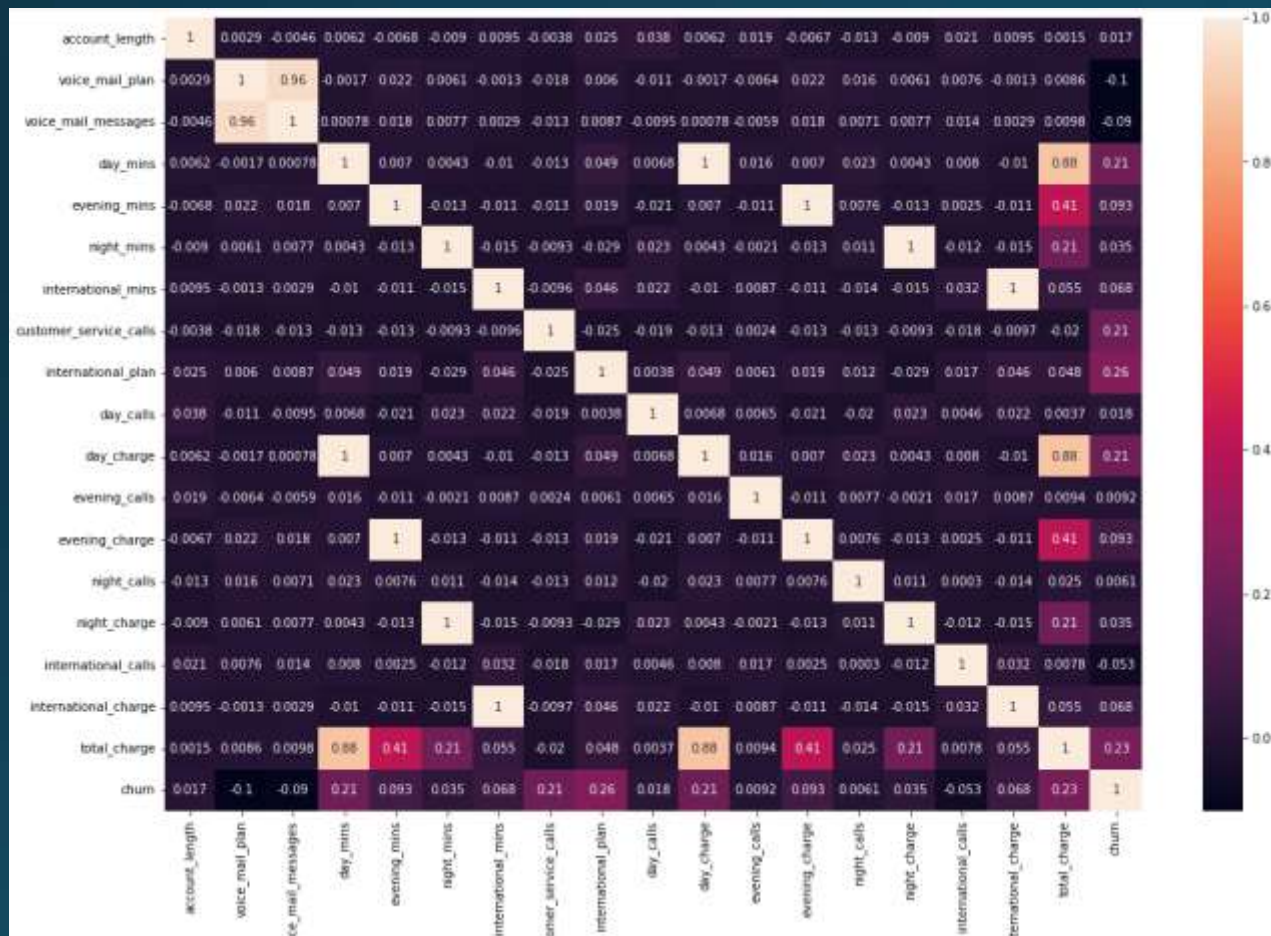
Summary of the statistical analysis

- The average time customers talk during day, evening, night is 198.88 min/day.
- The average time a customer talks on a international call is 10.24 min /day.
- In the daytime 75% of the customer talk on more than 200 min/day.
- Also we can clearly see that the average and percentile of evening and night minute are almost equal i.e. in both cases 75% of the data falls under 235.30 min/day.
- Average cost per call for day is 3.29/call.
- Average cost per call for evening is 5.86/call.
- Average cost per call for night is 11.03/call.
- From the above we can conclude that average cost/call varies vastly depending on the time of the day.
- Average total charge is 59.45, maximum going up to 96.15

Various visualizations



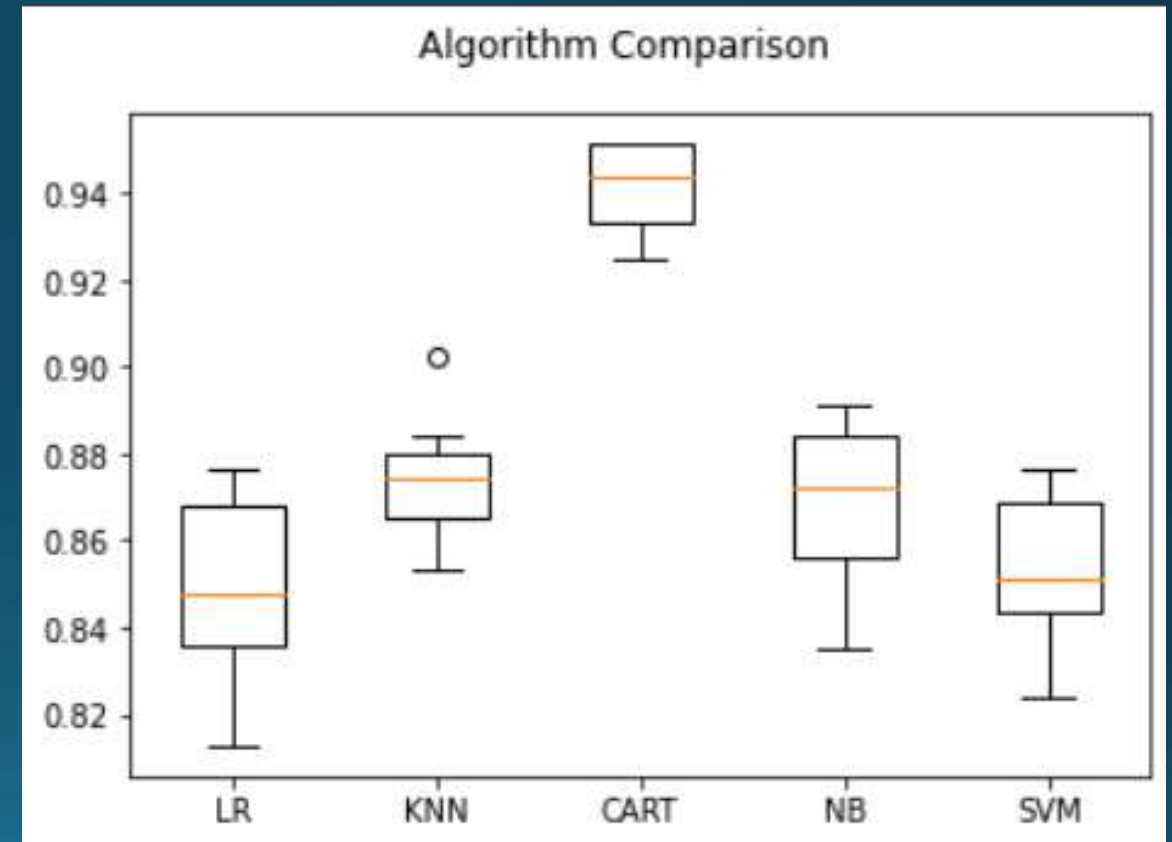
Correlation analysis



- Voice_mail_plan and voice_mail_messages are highly correlated.
- All the charges and min for day, evening and night are highly correlated to each other.
- This tells us that both the features moves in positive direction, if min on call increases the charge will also increase.
- Total charge is correlated with day min and charges
- This shows that total charge is mostly effected by the day charge. So if the day min and charge will increase the total charge will also increase.
- Except for these factors no other features are correlated to one another.

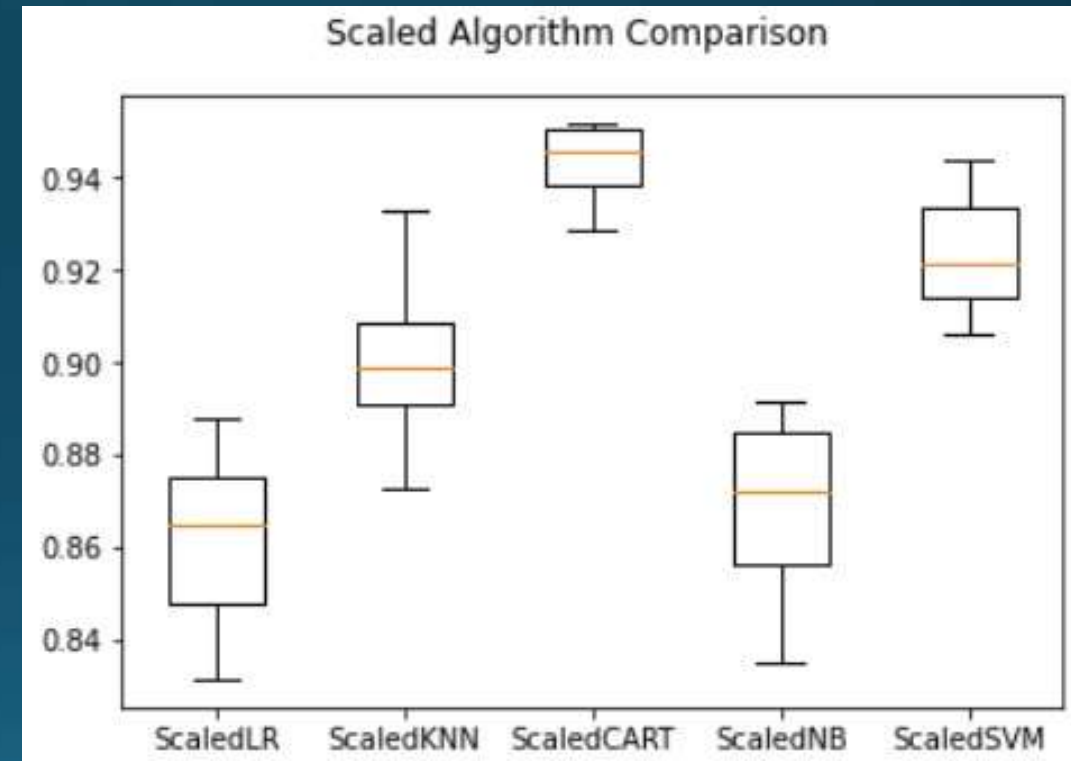
Deciding best algorithm

LR: 0.849197 (0.020277)
KNN: 0.874701 (0.012970)
CART: 0.941474 (0.009883)
NB: 0.869095 (0.018294)
SVM: 0.853702 (0.016692)



Deciding best algorithm after standardizing the dataset

ScaledLR: 0.860833 (0.018150)
ScaledKNN: 0.899845 (0.018151)
ScaledCART: 0.943730 (0.007362)
ScaledNB: 0.869095 (0.018294)
ScaledSVM: 0.923098 (0.013483)

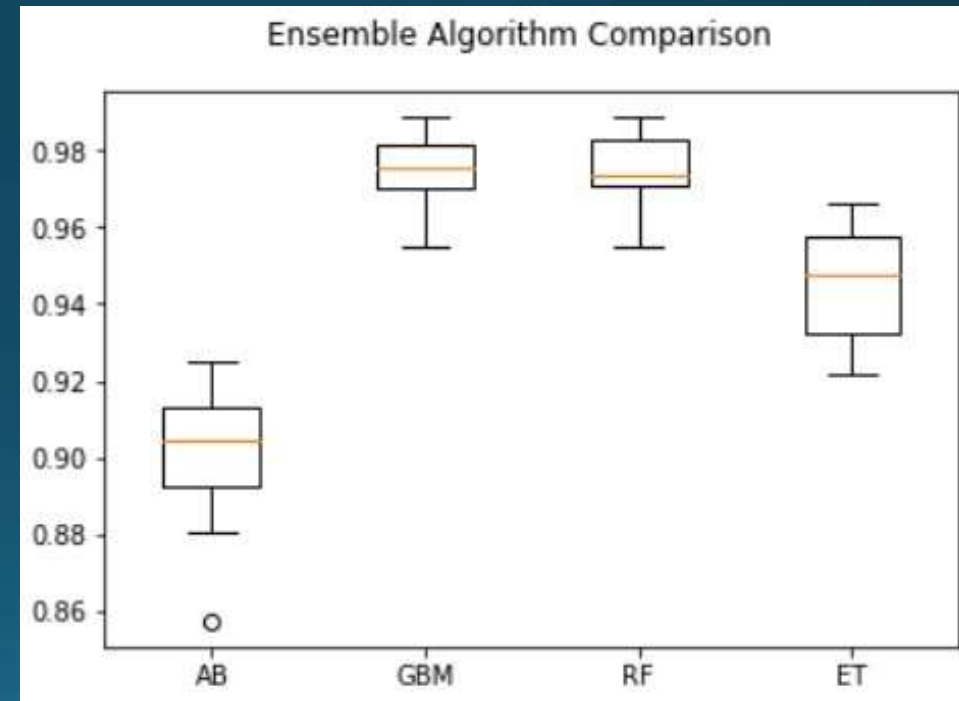


KNN - Algorithm

```
0.872459 (0.013170) with: {'n_neighbors': 1}  
0.897211 (0.014353) with: {'n_neighbors': 3}  
0.899842 (0.019582) with: {'n_neighbors': 5}  
0.902092 (0.018958) with: {'n_neighbors': 7}  
0.894969 (0.014834) with: {'n_neighbors': 9}  
0.891969 (0.014716) with: {'n_neighbors': 11}  
0.892343 (0.016020) with: {'n_neighbors': 13}  
0.891217 (0.016463) with: {'n_neighbors': 15}  
0.889717 (0.015696) with: {'n_neighbors': 17}  
0.888214 (0.014566) with: {'n_neighbors': 19}  
0.885212 (0.013412) with: {'n_neighbors': 21}
```


Ensemble Algorithm Comparison

AB: 0.899844 (0.019103)
GBM: 0.974860 (0.010369)
RF: 0.974860 (0.010368)
ET: 0.945613 (0.014833)



Final model – Random Forest

```
# load the model from disk
loaded_model = load(open('filename', 'rb'))

result = loaded_model.score(rescaledValidationX, Y_validation)
print(result)

0.9880059970014993
```

- We finally chose the Random forest algorithm for the model as the highest accuracy shown among all of the algorithms was this
- Accuracy of **98.8%**

Deployed model using Streamlit

User Input parameters

	account_length	voice_mail_plan	voice_mail_messages	day_mins	evening_mins	night_mins	internat
0	380.0000	1	56.0000	824.9900	22.0000	48.0000	

Predicted Result

Customer is Loyal

THANK YOU